

2013 年度 修士論文要旨

TBC の規則改良による少量学習データを用いた音響モデル構築

関西学院大学大学院理工学研究科

情報科学専攻 川端研究室 中村貴亮

大語彙連続音声認識の音響モデルは、母音や子音など単一の音素を 3 つ組にしたトライフォンを学習単位とし、統計的手法によってそれらの典型的特徴量を学習する。このとき、統計的学習の信頼性を高めるためには大量の音声データを必要とするが、トライフォンの種類数は非常に多いため全てのモデルに対して十分な学習データ量を用意することは難しい。

そこで、音声学的な知見による分類規則に従って、類似するトライフォンをグループ化し、グループ内のトライフォンが音響モデルを共有する TBC 法がよく用いられる。しかし、音声認識の対象を種々の言語や方言に拡大する過渡期においては、学習データが更に不足するため、TBC による認識精度が低下する恐れがある。この問題に対して、グループ化を決定する基準を改良したものは数多くあるが、分類規則に注目したものは少ない。そこで本論文では、少量の学習データによる音響モデルを構築するために、分類規則を精密化することを試みた。特に、摩擦音の体系や清音・濁音の区分に着目し、共有化する音素文脈を分割あるいは結合した。

日本語話し言葉コーパス (CSJ) の学会講演を学習データとして音響モデルを構築するとき、学習データの削減率が 70 % を超えると、単語誤り率の増加・クラスタ数の減少の度合いが大きくなり、木構造によるクラスタリングが適切でなくなる。よって、学習データが不十分な状況を再現するためには、CSJ の学習データの削減率を 70 % 以上にすべきであることが予備実験によって判明した。

十分な学習データを用いて構築した音響モデルの認識誤りの傾向から、分類規則を改定することによって単語誤り率が減少することを示した。具体的には、先行音素が歯茎摩擦音のとき、清音と濁音を区別するよう質問内容を分割し、後続音素が歯茎摩擦音もしくは硬口蓋摩擦音のとき、清音と濁音のみ区別するよう質問内容を結合した。また、データセットを複数用意し、少量の学習データによる話者依存の音響モデルを構築し、単語誤り率が減少したことから、改定した分類規則が一般性を持つことを示した。

CSJ の学習データを 70, 80, 90 % 削減することによって学習データが不足する状況を再現し、少量の学習データによる話者独立の音響モデルを構築した。その結果、一部を除く全ての削減率・混合数において単語誤り率が減少した。このことから、改定した分類規則に従うクラスタリングは有効に働き、クラスタの適切な初期値を得ることができたと考えられる。